REMARKS

In response to the Office Action mailed on March 5, 2009, the new Assignee (Nuance Communications, Inc.) respectfully requests reconsideration in view of the foregoing amendments and the following remarks. To further prosecution of this application, each of the rejections set forth in the Office Action has been carefully considered and is addressed below. Additionally, each of the independent claims has been amended to make explicit some aspects of the claims that are believed to have been implicit in the claims. No new matter has been added. The application is believed to be in condition for allowance.

Claim Rejections – 35 U.S.C. §101

Claims 1-5 and 17-20 stand rejected under 35 U.S.C. §101 as purportedly not falling within one of the four statutory categories of invention. (Office Action, page 5). More particularly, independent claims 1 and 17 purportedly recite purely mental steps and would not qualify as a statutory process.

Without acceding to the propriety of the rejection, claims 1-5 and 17-20 have been amended in a manner that is believed to overcome this rejection. Claim 1 now recites at least one computer readable medium encoded with instructions that, when executed by at least one processor, perform a method for generating speech recognition models. Claim 17 now recites at least one computer readable medium encoded with instructions that, when executed by at least one processor, perform a method for recognizing speech from an audio stream originating from one of a plurality of data classes. Support for these amendments are found at least on page 5, lines 15-18 of the specification. As amended, claims 1 and 17 are now directed to a particular machine or manufacture in the form of a computer readable medium. Accordingly, the rejection of claims 1-5 and 17-20 under §101 should be withdrawn.

Claim Rejections - 35 U.S.C. §103

Each of independent claims 1, 6, 11 and 16 stands rejected under 35 U.S.C. §103(a) as purportedly being unpatentable over Chang (U.S. Patent No. 6,567,776) in view of Yang (U.S. Patent Publication No. 2001/0010039). Each of dependent claims 2-5, 7-10 and 12-15 has been

rejected as purportedly being unpatentable over the combination of Chang and Yang, and for some claims, in further view of Kanevsky (U.S. Patent No. 6,529,902).

Each of independent claims 17, 21 and 24 stands rejected under 35 U.S.C. §103(a) as purportedly being unpatentable over Wark (U.S. Patent Publication No. 2003/0231775) in view of Chang in further view of Yang. Each of dependent claims 18-20, 22-23 and 25-27 has been rejected as purportedly being unpatentable over the combination of Wark, Chang and Yang.

These rejections are respectfully traversed.

A. Overview of Embodiments of the Invention

Speech recognition is the process by which computers analyze sounds and attempt to characterize them as particular letters, words, or phrases. Generally, a speech recognition system is "trained" with many phoneme examples. (Page 1, lines 8-11). A speech recognition system examines various features from each phoneme example by mathematically modeling its sounds on a multidimensional landscape using multiple Gaussian distributions. (Page 1, lines 18-20). Once acoustic models of phonemes are created, input speech to be recognized is sliced into small samples of sound that are each converted into a multidimensional feature vector by analyzing the same features as previously used to examine the phonemes. (Page 1, lines 21-24). Speech recognition is then performed by statistically matching the feature vector with the closest phoneme model, wherein the accuracy, or word error rate, of a speech recognition system is dependent on how well the acoustic models of phonemes represent the sound samples input by the system. (Page 1, lines 24-29).

Gender specific models, i.e., separate female and male acoustic models of phonemes, are known to yield improved recognition accuracy over gender independent models. (Page 1, line 30 to page 2, line 1). The conventional use of such models is to build one system with just female models and one system with just male models, wherein samples are decoded using both systems in a two-pass approach. (Page 2, lines 1-4). While such gender specific systems provide better speech recognition results, Applicants indicate that they generally require too much computing power and resources to be practical in many real-world applications. (Page 2, lines 4-6).

Embodiments of the invention are directed to addressing such limitations of conventional speech recognition systems by generating efficient gender dependent models and integrating such models with an efficient class detection scheme. (Page 2, lines 8-11). Embodiments of the invention determine which models contain class independent information and create class independent models in place of such models. (Page 2, lines 11-13). Embodiments of the invention teach a highly accurate class detection scheme to detect class at a computational cost that is negligible. (Page 2, lines 13-15).

One embodiment of the invention is directed to a technique for generating recognition models. As illustrated in FIG. 1, female training data 104 and male training data 106 is received that contains thousands of recorded phonemes spoken by male and female speakers, where each training data is identified by its phoneme and whether it comes from a male speaker or a female speaker. (Page 3, line 27 to page 4, line 3). The phonemes in the female and male training data are modeled by quantifying various features from the data, including the data's signal frequencies, intensities, and other characteristics. (Page 4, lines 6-9). Female models 110 are created based solely on the female training data 104 and male models 112 are created based solely on the male training data 106. (Page 4, lines 13-15).

Each female model 110 and male model 112 is compared for each phoneme to determine if the gender separation is insignificant. (Page 4, lines 16-18). For female and male phonemes that are insignificantly different from each other, their female and male training data 104 and 106 are combined and gender independent (GI) models 114 are created. (Page 4, lines 28-31). Additionally, the separate female models 110 and male models 112 that are determined to have insignificant differences from one another are removed. (Page 4, line 31 to page 5, line 1). This results in female models 110 derived from female training data 104, male models 112 derived from male training data 106, and gender independent models 114 derived from both the female and male training data 104 and 106, wherein the female models 110 and male models 112 are significantly different from each other. (Page 5, lines 1-6).

The model creation system and technique beneficially reduces the amount of acoustic models needed to be stored and searched during speech recognition. (Page 5, lines 7-9). Furthermore, a speech recognition system using the female, male and gender independent models

created using this technique requires less computing power, uses less system resources, and is more practical to implement with minimal loss in recognition accuracy. (Page 5, lines 9-13).

FIG. 2 illustrates one process for generating Gaussian Mixture Models (GMMs) according to one embodiment of the invention. The process begins by creating gender independent models (GMM_{GI}) from both female and male training data during a training operation 202. (Page 5, lines 24-26). Additionally, female models (GMM_f) are created and trained from just the female training data, and male models (GMM_m) are created and trained using just the male training data, during training operations 204 and 206. (Page 6, line 27 to page 7, line 1). The various models can be created with training operations that may be performed in sequence or in parallel. After creating and training the models, the female models are compared with the male models to determine if their differences are insignificant, for example, by measuring the differences using the Kullback Leibler divergence. (Page 7, lines 5-10).

For those phonemes which are determined to carry insignificant gender information, the gender independent models for these phonemes are added to a final system model. (Page 9, lines 8-12). For those phonemes that carry gender information determined to be significant, separate female models and male models for phonemes with significant gender information are added to the final system model. (Page 9, lines 13-17). The process continues until examination of all the phoneme models is completed.

In another embodiment, a process for speech recognition that employs the generated female models, male models and gender independent models is illustrated in FIG. 4. More particularly, this aspect of the invention involves a method for recognizing data from a data stream originating from one of a plurality of data classes that includes the female, male and gender independent models described above.

It should be appreciated that the foregoing discussion of embodiments of the invention is provided merely to assist the Examiner in appreciating various aspects of the present invention. However, not all of the description provided above necessarily applies to each of the independent claims pending in the application. Therefore, the Examiner is requested to not rely upon the foregoing summary in interpreting any of the claims or in determining whether they patentably

distinguish over the prior art of record, but rather is requested to rely only upon the language of the claims themselves and the arguments specifically related thereto provided below.

B. Each of the Independent Claims Distinguishes Over the Applied References

Without acceding to the propriety of the rejections, each of the independent claims has been amended to explicitly set forth aspects of the invention that are believed to have been implicit in the claims and distinguish over the combination of Chang and Yang, and the combination of Wark, Chang and Yang.

Independent claims 1, 6, 11 and 16

As amended, independent claim 1 recites determining a difference in model information between pairs of corresponding phoneme models of the female speech recognition model and the male speech recognition model; and creating a gender-independent speech recognition model that includes a gender-independent phoneme model based on a pair of corresponding phoneme models of the female speech recognition model and the male speech recognition model when the difference in model information between the phoneme models of the pair of corresponding phoneme models is insignificant.

Independent claims 6, 11 and 16 have been amended in a similar manner to explicitly recite the aspect of creating an independent speech recognition model that includes an independent phoneme model based on a pair of corresponding phoneme models of first and second speech recognition models when the difference in model information between the phoneme models of the pair of corresponding phoneme models is insignificant.

In the Office Action, it is contended that Chang discloses a method for generating speech recognition models that involves each of the acts recited in claims 1, 6, 11 and 16, except for the use of phonemes training data and creating a gender-independent speech recognition model based on a female set of recorded phonemes training data and a male set of recorded phonemes training data if the difference in model information is insignificant. (Office Action, pages 6-7). To cure this acknowledged deficiency, the Office Action looks to Yang for the proposition that it purportedly teaches well known techniques of speech recognition, wherein differences are evaluated between all

voice types. (Office Action, page 7). It is contended that it would purportedly have been obvious to one of ordinary skill in the art at the time of the invention to modify Chang to incorporate phoneme training data and create a gender-independent speech recognition model based on a first set of recorded phonemes training data and a second set of recorded phonemes training data if the difference in model information is insignificant, as purportedly taught by Yang. (Office Action, page 9). The Assignee respectfully disagrees.

Chang discloses a speech recognition method that uses speaker cluster models. Chang describes that it was known to apply speaker cluster models to speaker-independent speech recognition and speaker adaptation. (Col. 1, lines 14-16). The speaker cluster models are built in the same training phases that start with dividing speakers into different speaker clusters, and then independently training a cluster-dependent model for each speaker cluster using the speech data of the speakers belonging to the cluster. (Col. 1, lines 18-22). The collection of all the cluster-dependent models then forms a speaker cluster model. (Col. 1, lines 22-23). Chang indicates that most approaches in building speaker cluster models were focused on means of dividing speakers into clusters, especially in finding measurement of similarities across speakers. (Col. 1, lines 23-26).

In the training phase of the speaker cluster model, the known methods emphasized how to cluster speakers with similar characteristics into the same speaker cluster. (Col. 2, lines 20-23). However, Chang indicates that improving the effectiveness of speaker clustering does not necessarily improve the accuracy of speech recognition because each cluster-dependent model is viewed as an independent recognition model during a recognition phase, and the dependency among different cluster-dependent models is never considered. (Col. 2, lines 23-32). Chang sought to improve the performance of speech recognition by introducing the dependency among a plurality of cluster-dependent models to overcome recognition problems caused by between-speaker variability. (Col. 2, lines 60-67).

Chang employs a tree-structured speaker cluster model, as shown in FIG. 1, having multiple levels and a plurality of speaker clusters. The model includes a root speaker cluster 100 that uses all of the speech data to train a speaker-independent model. (Col. 5, lines 1-4). All speakers are then clustered according to gender into a male speaker cluster 102 and a female speaker cluster 104 that

are used to train a gender-dependent model. (Col. 5, lines 4-6). The speakers within each gender group are then clustered into separate male speaker clusters 112, 114 and female speaker clusters 122, 124. In training the speaker cluster model, Chang employs a discriminate function during the training phase.

As one of ordinary skill in the art would readily appreciate, nowhere does Chang teach, suggest or otherwise recognize the creation or use of an independent speech recognition model that includes an independent phoneme model based on a pair of corresponding phoneme models of first and second speech recognition models when the difference in model information between the phoneme models of the pair of corresponding phoneme models is insignificant. Rather, to the extent that Chang discloses an independent-speaker model, it forms the root level of a tree-structured speaker cluster model which includes all the speech data for training the model and the lower level speaker-dependent models. In other words, the Chang independent speaker model is used to create various speaker dependent clusters; it is not created from other speech recognition models when the difference in model information between the phoneme models of a pair of corresponding phoneme models of other speech recognition models is insignificant.

Yang fails to cure the deficiencies of Chang. More particularly, Yang is directed to Mandarin Chinese speech recognition by using initial/final phoneme similarity vector. Like Chang, Yang also fails to teach, suggest or even recognize the creation of an independent speech recognition model that includes an independent phoneme model based on a pair of corresponding phoneme models of first and second speech recognition models when the difference in model information between the phoneme models of the pair of corresponding phoneme models is insignificant.

In view of the foregoing, independent claims 1, 6, 11 and 16 patentably distinguish over Chang and Yang, taken either alone or together, which fail to teach or suggest each limitation of the claims. Accordingly, the rejection of independent claims 1, 6, 11 and 16 under §103 as being obvious in view of Chang and Yang should be withdrawn.

Independent claims 17, 21 and 24

As amended, independent claim 17 recites a gender-independent speech recognition model that includes independent phoneme models based on pairs of corresponding recorded phonemes originating from the plurality of female speakers and the plurality of male speakers having insignificant differences in model information between the recorded phonemes of the pair of corresponding recorded phonemes. Claim 17 also recites that each of the female speech recognition model and the male speech recognition model lacks the phoneme models of the gender-independent speech recognition model based on pairs of corresponding recorded phonemes originating from the plurality of female speakers and the plurality of male speakers having insignificant differences in model information between the recorded phonemes of pairs of corresponding recorded phonemes.

As amended, independent claims 21 and 24 recite a third speech recognition model that includes phoneme models based on pairs of corresponding recorded phonemes originating from both the first and second set of speakers having insignificant differences in model information between the recorded phonemes of the pair of corresponding recorded phonemes. The claims also recite that each of the first speech recognition model and the second speech recognition model lacks the phoneme models of the third speech recognition model based on pairs of corresponding recorded phonemes originating from both the first and second set of speakers having insignificant differences in model information between the recorded phonemes of the pairs of corresponding recorded phonemes.

The Office Action contends that Wark teaches a system for recognizing speech data from an audio stream originating from one of a plurality of data classes and includes each of the features recited in independent claims 17, 21 and 24, except for the particular data classes recited in the claims. (Office Action, pages 13-14). In an effort to cure this deficiency, the Office Action also relies upon Chang and Yang. (Office Action, pages 14-17).

Without acceding to either the characterization of Wark set forth in the Office Action or the propriety of the purported combination of references, Chang and Yang do not cure the deficiencies of Wark. As discussed above, Chang and Yang, either alone or together, do not teach or suggest speech recognition models as recited in independent claim 17, 21 and 24. More particularly, the references fail to teach or suggest data classes that include first (female) and second (male) speech

recognition models based on recorded phonemes originating from a first set of speakers (female) and a second set of speakers (male), and a third speech recognition model (gender-independent) that includes phoneme models based on pairs of corresponding recorded phonemes originating from the first and second sets of speakers having insignificant differences in model information between the recorded phonemes of the pair of corresponding recorded phonemes, and wherein each of the first and second recognition models lacks the phoneme models of the third speech recognition model based on pairs of corresponding recorded phonemes originating from the first and second sets of speakers having insignificant differences in model information between the recorded phonemes of pairs of corresponding recorded phonemes.

In view of the foregoing, independent claims 17, 21 and 24 patentably distinguish over Wark, Chang and Yang, taken either alone or together, which fail to teach or suggest each limitation of the claims. Accordingly, the rejection of independent claim 17, 21 and 24 under §103 as being obvious in view of Wark, Chang and Yang should be withdrawn.

Dependent claims

Each of the dependent claims depends from one of independent claims 1, 6, 11, 16, 17, 21 and 24 and is patentable for at least the same reasons. Thus, the rejection of each of the dependent claims should similarly be withdrawn.

Since each of the dependent claims depends from a base claim that is believed to be in condition for allowance, the Assignee believes that it is unnecessary at this time to argue the further distinguishing features of the dependent claims. However, the Assignee does not necessarily concur with the interpretation of the dependent claims set forth in the Office Action, nor does the Assignee concede that the prior art alleged to show the features in the dependent claims does so. Therefore, the Assignee reserves the right to specifically address the further patentability of the dependent claims in the future.

CONCLUSION

In view of the foregoing amendments and remarks, this application should now be in condition for allowance. A notice to this effect is respectfully requested. If the Examiner believes, after this amendment, that the application is not in condition for allowance, the Examiner is requested to call the undersigned at the telephone number listed below.

If this response is not considered timely filed and if a request for an extension of time is otherwise absent, the Assignee hereby requests any necessary extension of time. If there is a fee occasioned by this response, including an extension fee, the Director is hereby authorized to charge any deficiency or credit any overpayment in the fees filed, asserted to be filed or which should have been filed herewith to our Deposit Account No. 23/2825, under Docket No. N0484.70762US00.

Dated: 6/5/09

Respectfully submitted,
Nuance Communications, Inc.

By _____
James M. Hanifin, Jr.
Registration No.: 39,213
Richard F. Giunta
Registration No.: 36,149
WOLF, GREENFIELD & SACKS, P.C.
Federal Reserve Plaza
600 Atlantic Avenue
Boston, Massachusetts 02210-2206
617.646.8000